Klug, Julia; Gerich, Mara; Schmitz, Bernhard

Can teachers' diagnostic competence be fostered through training and the use of a diary?

Journal for educational research online 8 (2016) 3, S. 184-206

Journal for Educational Research Online
Journal für Bildungsforschung Online
Volume 8 (2016), No. 3, 184–206

Julia Klug, Mara Gerich & Bernhard Schmitz

Can teachers' diagnostic competence be fostered through training and the use of a diary?

Abstract

Diagnostics is an integral part of teaching. To date, investigations have been carried out into the accuracy of teachers' judgments. Recently, a model of teachers' diagnostic competence describing diagnosing as a three-dimensional process has been developed and empirically tested. A necessity for promoting diagnostic was among teachers with regard to students' learning behavior was shown. The aim of this study was to develop and evaluate a training program which utilizes this new perspective. Forty-seven secondary school teachers participated in the quasi-experimental study, split into a waiting control group (n = 17), who did not participate in the training program, but did take part in a pre- and posttest along with Experimental Group 1 (n = 15), who took part in the training program; and Experimental Group 2 (n = 15), who took part and worked on semi-standardized diaries in order to self-monitor their implementation of diagnostic actions for four weeks, returning 176 diaries. Pre- and posttest measures were combined with time-series data. Results showed that the training program enhanced teachers' diagnostic competence, especially when it came to actions before and during diagnosing. The diary proved to be an accurate instrument for measuring transfer, but had no additional intervention effect above and beyond the training program. As pressures to provide individualized support to students increase, the evaluated training program will prove to be helpful.

Keywords

Training program; Diagnostic competence; Process; Diary; Time-series analysis

Dipl.-Psych. Dr. Julia Klug (corresponding author), Faculty of Psychology, Department of Applied Psychology: Work, Education and Economy, University of Vienna, Universitätsstraße 7 (NIG), 1010 Vienna, Austria
e-mail: Julia.Klug@univie.ac.at

Dipl.-Psych. Dr. Mara Gerich · Prof. Dr. Bernhard Schmitz, Institute of Psychology, Department of Educational Psychology, TU Darmstadt, Alexanderstraße 10, 64283 Darmstadt, Germany
e-mail: gerich@psychologie.tu-darmstadt.de
schmitz@psychologie.tu-darmstadt.de

Kann die diagnostische Kompetenz von Gymnasiallehrkräften durch ein Training und den Einsatz eines Tagebuchs gefördert werden?

Zusammenfassung

Diagnostizieren zählt zu den zentralsten Kompetenzen von Lehrkräften. Bisher wurde die diagnostische Kompetenz meist als Urteilsgenauigkeit operationalisiert. Kürzlich wurde ein Prozessmodell entwickelt und empirisch überprüft, das den Diagnoseprozess beschreibt, den Lehrkräfte zum Ziel der Diagnostik und Förderung durchlaufen. Es konnte gezeigt werden, dass es noch großen Förderbedarf gibt. Zum Ziel der Förderung nahmen 47 Gymnasiallehrkräfte an der vorliegenden Studie teil, die drei quasi-experimentellen Bedingungen zugeteilt waren: Einer Wartekontrollgruppe (n = 17), die nicht am Trainingsprogramm, aber wie die anderen beiden Gruppen auch am Prä- und Posttest teilnahm, Versuchsgruppe 1 (n = 15), die das Trainingsprogramm bekam und Versuchsgruppe 2 (n = 15), die zusätzlich über vier Wochen halbstandardisierte Tagebücher bearbeitete, um die Umsetzung ihrer diagnostischen Handlungen im Unterricht zu überwachen. Zur Evaluation wurden sowohl Prä-Post- als auch Zeitreihenanalysen durchgeführt. Die Ergebnisse zeigen, dass durch das Trainingsprogram insbesondere die Kompetenzen von Lehrkräften vor und während des Diagnostizierens gesteigert werden konnten. Das Tagebuch erwies sich als hilfreiches Instrument um den Transfer in den Unterricht zu messen. Allerdings zeigte sich kein zusätzlicher Interventionseffekt, der über den des Trainingsprogramms hinausgeht. Trainingsprogramms und Tagebuch können als hilfreiche Tools in der Lehrkräfteaus- und -weiterbildung verwendet werden.

Schlagworte

Trainingsprogramm; Diagnostische Kompetenz; Prozess; Tagebuch; Zeitreihenanalyse

1. Introduction

In going about their daily duties, teachers are faced with highly complex work and must often multitask (Brante, 2009). In addition to having professional knowledge, giving learner-centered instruction, managing their classes, interacting with students, and being motivated role models (Opdenakker & Van Damme, 2006; Kukla-Acevedo, 2009), teachers' key tasks include making diagnoses, as they are challenged to meet students' diverse learning needs and adapting their teaching to students with heterogeneous academic abilities as well as multiple interests and motivations (Vogt & Rogalla, 2009). Prior empirical research from the 1970s until today has focused particularly on investigating students' academic achievement. So far, accuracy in teachers' judgments has been operationalized as their ability to

accurately judge the difficulty of tasks or their students' performance. Accuracy has been measured by correlating teachers' judgments with the results of standardized tests (e.g., Coladarci, 1986; Feinberg & Shapiro, 2003; Helmke & Schrader, 1987; Lee, Chiu, van Hasselt, & Tong, 2009; Wang, 1973; Spinath, 2005). In this approach to the assessment of teachers' diagnostic competence, diagnoses of student achievement have been the focus. However, there is an ongoing push in the theoretical literature to shift the focus to diagnoses of learning behavior, which allows for didactic action afterwards, such as giving students individualized support and adapting classes to their needs (e.g., Abs, 2007). Recently, Klug, Bruder, Kelava, Spiel, and Schmitz (2013) developed and empirically tested a model of teachers' diagnostic competence that accounts for learning behavior, thus closing the gap between previous empirical research and recent theoretical demands. It describes the diagnosis of learning behavior as a three-dimensional process, consisting of a preactional, an actional, and a postactional phase. In a study based on that model, a high need for promoting diagnostic competence among teachers with regard to students' learning behavior was shown (Klug, Bruder, & Schmitz, 2016).

There is a growing awareness of the necessity of assisting teachers in their professional development in general (Bakkenes, Vermunt, & Wubbles, 2010). In particular, there is a call for further educational programs to cultivate new facets of teachers' diagnostic competence (Klieme et al., 2003). Nevertheless, only few such programs exist as of yet, and, more particularly, none of them consider the process approach accounting for students' learning behavior as modeled by Klug et al. (2013). Therefore, one purpose of this study was to develop and evaluate a training program based on the model of teachers' diagnostic competence by Klug et al. (2013) in order to promote teachers' diagnostic competence with regard to learning behavior.

Research on how to gain expertise in a field and various models of teachers' professional development have shown that reflection on action is a crucial variable for developing competencies (e.g., Bakkenes et al., 2010; Berliner, 2001; Clarke & Hollingsworth, 2002; Epstein & Hundert, 2002; Marcos, Miguel, & Tillema, 2009; Sowa, 2009; Strasser & Gruber, 2003). The use of learning diaries as a method to reflect upon one's actions has been successfully applied among pupils at schools and university students (e.g., Schmitz, Klug, & Schmidt, 2011). Diaries force one to continuously record and reflect on one's own behavior, and can lead to changes in behavior in a desired direction (Webber, Scheuermann, McCall, & Coleman, 1993). Thus, the second purpose of our study was to develop a diary for teachers to reflect upon their diagnostic behavior, and to put it into practice. Working with the diary on a regular basis should (a) bolster teachers' transfer of the learned content from the training program, and (b) allow us to measure the transfer to teachers' classroom actions.

2. Theory

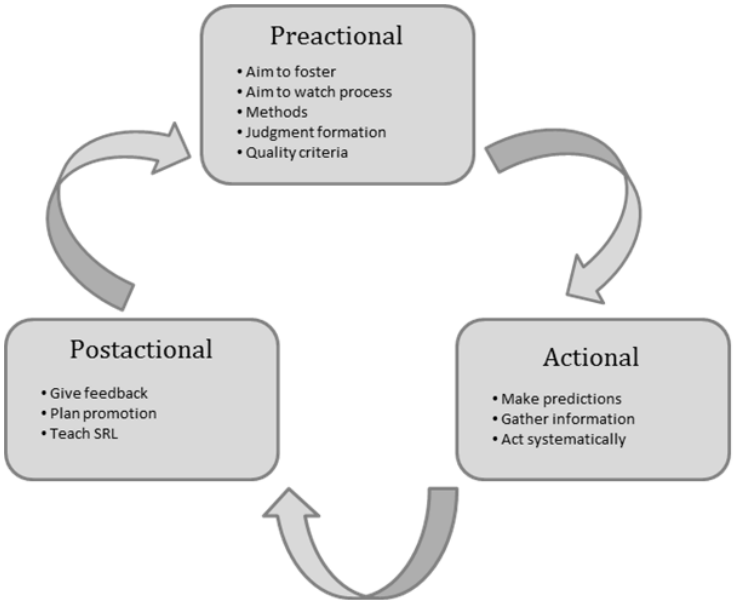
We decided to base both our training program and the diary on the newly developed process model of diagnostic competence that accounts for the cross-curricular diagnosis of learning behavior (Klug et al., 2013). This model will now be described in advance of the subject of diaries and their advantages and effects.

2.1 A Process model of teachers diagnostic competence that accounts for diagnosing learning behavior

The model's domain is teachers' diagnoses. It is context-specific in focusing on diagnoses of pupils' learning behavior, both at school and while learning at home. It addresses questions such as how pupils approach learning tasks, how they do their homework, how they study at home, which learning strategies they are able to apply, and how much they self-regulate their learning. According to Wirth and Leutner (2008), self-regulated learning (SRL) is "a learner's competence to autonomously plan, execute, and evaluate learning processes, which involves continuous decisions on cognitive, motivational, and behavioral aspects of the cyclic process of learning" (p. 103). It includes students' metacognitive strategies for planning, monitoring, and modifying their cognition, students' management and control of their effort on classroom academic tasks, and students' actual cognitive strategies which they use to learn, remember, and understand the material (Pintrich & van De Groot, 1990). Several correlational studies (e.g., Pintrich & van De Groot, 1990; Yen, Konold, & McDermott, 2004), training studies (e.g., Perels, Dignath, & Schmitz, 2009) and meta-analyses (Dignath, Büttner, & Langfeld, 2008; Hattie, Biggs, & Purdie, 1996) support the connection between students' self-regulated learning behavior and their academic achievement. Teachers can promote students' self-regulated learning if they diagnose problems in this area (e.g., Klug, Krause, Schober, Finsterwald, & Spiel, 2014; Perry, Hutchinson, & Thauberger, 2008). This can in turn lead to better long term achievement. We understand teachers' diagnostic competence of self-regulated learning as well as self-regulated learning itself as generic competences that are important in, and can be transferred to, every school subject.

In the model of teachers' diagnostic competence, diagnosing is conceptualized as a process (Jäger, 2007). In the style of process models of self-regulation consisting of three cyclical phases – before, whilst and after learning (Schmitz & Wiese, 2006; Zimmerman, 2000) –, the model consists of three cyclical dimensions for the diagnosis process, which take place in a preactional, an actional, and a postactional phase. The factor structure has been empirically tested, and the model fit the empirical data well (Klug et al., 2013). Figure 1 illustrates the process model for diagnosing learning behavior.

Figure 1: Process model of teachers' diagnostic competence concerning pupils' learning behavior (Klug et al., 2013)



In the following section, the three phases of the diagnosis process will be described.

2.1.1 Preactional phase

In the preactional phase, every diagnosing action before summarizing the information into an actual diagnosis of a pupil's learning behavior matters. The teacher needs to specify the target of the diagnosis. For example, observing one individual student's learning process for a specific topic and providing support to the student on the basis of this diagnosis (Horstkemper, 2004; Abs, 2007). Focusing on each student's unique learning processes is especially important in helping teachers develop an individual frame of reference, instead of a social one. This reduces the big-fish-little-pond effect – the notion that high-achieving pupils have a lower self-concept in a class of high-achieving peers, and vice versa (Lüdtke, Köller, Marsh, & Trautwein, 2005). Furthermore, in the preactional phase, the teacher's basic diagnostic skills are activated (Strasser & Gruber, 2003). These include knowledge of methods of gathering information (Helmke, Hosenfeld, & Schrader, 2004), quality criteria of tests, and judgment formation (Van Ophuysen, 2006). The teacher should be familiar with methods of gathering information, should know how to use them, and should know which method is most effective in which situation, and whether the methods meet quality criteria as teachers need to be able to rationally analyze their assessment practices (Maclellan, 2004). Judgment biases, a construct from social psychology, can also influence teachers' diagnoses. Fiedler, Walther,

Freytag, and Plessner (2002) investigated the influence of these biases in a simulated classroom. It is important for teachers to be aware of these biases before making a diagnosis in order to avoid them while diagnosing.

2.1.2 Actional phase

In the actional phase, the actual diagnostic action takes place. The teacher's action must be systematic in order to get a reliable diagnosis, similar to following the scientific method when doing quantitative research (Wilson, 1952). This systematic action should begin with making a hypothesis about a student's development and possible underlying learning difficulties. This is similar to practicing medicine, where clinicians can use clinical prediction rules (McGinn, Jervis, Wisnivesky, Keitz, & Wyer, 2008). After that, the teacher should gather information from different sources and evaluate its relevance before finally interpreting the data and coming to a concluding diagnosis. Finally, the teacher can compare real developments with predicted ones in order to possibly make changes to their modus operandi for subsequent diagnoses.

2.1.3 Postactional phase

The postactional phase begins right after a diagnosis has been made, when the pedagogical action that follows from the diagnosis should be implemented (Abs, 2007). This includes giving feedback to students and parents. Feedback has been found to significantly influence students' self-regulated learning when given in an effective way (Butler & Winne, 1995; Hattie & Timperley, 2007). Diagnosing also builds the basis for the appropriate counseling of parents (Klug, Bruder, Keller, & Schmitz, 2012). Additionally, it is important to write plans for promoting the individual student's competencies. These plans should contain the skills that need to be worked on, the student's current skill level, the goals to be reached, and the methods by which these goals are intended to be reached. Finally, adapting a class in reaction to the diagnosis by teaching appropriate learning strategies and self-regulated learning can lead to even better academic performance (Pintrich & van De Groot, 1990).

As can be seen in Figure 1, the model has a cyclical character. The three phases proceed in chronological order and influence each other. Furthermore, a basic assumption is that one diagnosis situation influences consecutive diagnosis situations. Our training program reflects the content of these three cyclical phases.

2.2 Diaries

According to Webber et al. (1993), continuously recording one's behavior can lead to modifications of behavior in a desired direction. Diaries can be used to keep

continuous track of one's own behavior. The mechanism that produces the change in behavior is self-monitoring. Lan (1996) defines self-monitoring as the deliberate focusing of attention on one specific aspect of one's own behavior. Schmitz and Perels (2011) were able to show that students who worked on a learning diary achieved better results in a mathematical problem solving test than a control group. Their self-regulation and self-efficacy scores also increased.

Diaries can be constructed with different levels of standardization. For evaluation purposes, a high degree of standardization is suggested, whereas questions with an open answering format more intensely encourage reflection. In contrast to typical questionnaires which measure traits, diaries measure states. The value of a specific variable at the current moment is measured at regular intervals over a defined period of time. With regards to measurement, diaries are a very appealing and useful tool. As they are collected in the field immediately following the behavior of interest, their ecological validity is exceptionally high (Panadero, Klug, & Järvelä, 2016; Schmitz, 2006). Another big advantage of diaries is the many possible ways of diary data analysis. It is possible to examine intraindividual processes over time (Schmitz, 2006). Possible methods of analysis include, for example, testing whether the course of a variable follows a linear trend, or testing the effects of an intervention by means of interrupted time-series analysis (Campbell & Stanley, 1963).

Diaries can promote learning through three different mechanisms: (a) They can increase the effectiveness of interventions by promoting the transfer of the learned content. (b) They can promote self-regulation by activating monitoring processes. (c) They can enhance motivation, since even the smallest progresses are visible (Schmitz, Klug, & Schmidt, 2011).

In this study, we developed a diary for teachers, as opposed to schoolchildren or university students, who have been the target groups of most diary studies so far. For our diary, we chose a semi-standardized approach. The diary primarily contained standardized questions on every phase of the diagnostic process from the aforementioned model. Additionally, teachers had the possibility to reflect upon their daily diagnostic action and which training content they actually applied in the classroom. The mechanisms we expected to be at work in this case were the promotion of transfer, as well as the motivation to apply the training content.

2.3 Aims

Based on the theoretical assumptions of diagnostic competence concerning learning behavior, we developed a training program and a semi-standardized diary for teachers to bolster their diagnostic competence. Our hypotheses are as follows:

1) We expect an increase in teachers' diagnostic competence behavior if they participate in the training program in contrast to a control group. The increase is expected to be observable in the overall score, as well as in the scores for each phase of the diagnostic process and the corresponding variables.

2) For teachers who work on the diary, we expect an intervention effect in addition to the one from the training program because of the supplementary self-monitoring.
3) In the process data collected in the diaries, we expect positive linear trends for each trained variable across the training period, meaning that teachers who work on the diary are actually applying the trained methods more and more in their classes, therefore indicating that the transfer is succeeding.
4) Finally, in the diary data, we expect a lasting augmentation of scores for each training variable from a baseline to just after the session in which the specific variable was trained.

3. Method

3.1 Participants

In our study, 47 secondary school teachers from two German secondary schools (one *Gymnasium* or grammar school and one comprehensive school), and one teacher training college in the federal states of Hessen and Baden-Württemberg participated. Their mean age was 40.4 years (min. = 23, max. = 61) and their mean school-teaching experience was 9.89 years (min. = 1, max. = 38). Thirty-two (68 %) were female; 15 had their core teaching areas in mathematics or natural sciences, and 32 in languages or social sciences. Seven had already taken part in a further education program on diagnostics. For all teachers, participation was voluntary. As an incentive, teachers were given the opportunity to get further education credit points and a voucher for a book.

3.2 Design

The longitudinal, quasi-experimental design combined pre- and posttest measures with time-series data. There were two experimental groups and one control group. Experimental Group 1 (EG 1; n = 15) completed the pretest, then had three weekly training sessions, and completed the posttest afterwards. Experimental Group 2 (EG 2; n = 15) additionally worked on a semi-standardized diary, starting one week before the first training session and finishing one week after the last session. The control group (n = 17) participated in the pretest and posttest and were offered the opportunity to enroll in a shortened training program afterwards. The sample size seems small at first glance. However, for an ANOVA with repeated measures and a within-between interaction, the optimal total sample size we defined a priori using GPower was N = 47 (for three groups, two occasions of measurement, α = 0.01, medium estimated effect size, and r = .3 between the repeated measures).

3.3 Procedure

In the first session, a pretest containing a test of teachers' diagnostic competence, and a questionnaire with some demographic data was given. After that, the training program began. A similar test was given at the end of the last training session, supplemented by an evaluation of the training program on the reaction level. The training program took place in three weekly 180-minute after-class sessions in each of the schools. To ensure good training conditions, the teachers were trained in three subgroups of no more than 10 participants. EG 2 additionally worked on a semi-standardized diary in order to self-monitor their diagnostic actions for four weeks. We expected the self-monitoring to support the training transfer.

3.3.1 Training program

The training program was developed to foster teachers' diagnostic competence and covers the three phases of the diagnostic process in the model. Table 1 summarizes the content of each training session in relation to the phases of the model.

Table 1: Content of the training sessions in relation to the phases of the model

Session	Phase in the model	Content
1	Preactional & actional phase	Pretest Become acquainted with each other **Process of diagnosing, the systematic approach** Self-assessment Own specific case **Judgment formation** Reflection Homework
2	Preactional & actional phase	Review **Setting aims** **Making predictions** **Gathering information** **Methods** **Quality criteria** Reflection Homework
3	Postactional phase	Review **Teaching SRL** **Planning promotion** **Giving feedback to students and parents** Reflection Posttest

Note. The variables contained in the model are printed in bold letters.

In terms of didactics, each session contained a great deal of activity and reflection. Most importantly, participants worked on a specific case relating to one of their own students whom they had chosen in the first session. Additionally, participants were required to do homework for the following session. The content of

the previous sessions was reviewed at the beginning of each consecutive session. Furthermore, the trainer made clear that teachers were the experts at their schools and with their classes, and that with the training program, teachers would be given several methods to choose from. They could decide which methods would best facilitate their diagnosing action.

3.4 Instruments

3.4.1 Measurements of the pretest-posttest evaluation

For the pretest and posttest, a scenario test with open-ended questions to measure diagnostic competence on the basis of the model was chosen (Klug et al., 2013). Furthermore, some demographic data was collected with a short questionnaire. The pretest and posttest were conducted within the sessions shortly before the intervention started and after it ended.

Case Scenario. With a scenario test, we measured diagnostic competence concerning learning behavior based on the model, using a method as close to assessing real action as efficiently possible. The test consists of a case description of a pupil with certain difficulties in self-regulated learning, leading to a decline in his school performance. The teacher is asked to imagine being this student's teacher. The case description is followed by 12 open-ended questions, which are formulated with respect to the content of the model. Answers to each question are rated from 0 to 3 points. Total possible scores in the case scenario range from 0 to 36 as a sum score of the ratings from each answer. Possible scores for each of the three phases of the model (preactional, actional, postactional) range from 0 to 3 points as a mean score of the ratings of all answers corresponding to that phase. Raters receive a handbook with detailed references about how to rate the answers. For more information on the case scenario, see Klug et al. (2013). In this study, the scenario test proved to be valid for predicting an adequate diagnosis. Inter-rater reliabilities on each question were good with all values between ICC = .67 and ICC = .95.

3.4.2 Measurement of the process evaluation

The process evaluation was based on the standardized items in the diagnosis diaries given to teachers in EG 2. The diary was newly constructed for this study. It was also constructed with reference to the process model of diagnostic competence. The items mainly reflect the content of the model (11 items). Additionally, there are two introductory items on the teacher's current mood and the current day in class, and a final question on the teacher's satisfaction with his or her own diagnostic action that day. The introductory questions and the final question are answered on a scale using five smileys to show the range of effects. These items are not analyzed. The items related to the model are answered on a 6-point Likert scale ranging from

1 (*strongly disagree*) to 6 (*strongly agree*). Item 3 is inverted to control response sets. At the end of the diary, there is an additional open-ended question asking the teacher which techniques he or she was able to apply easily in class that day, and how he or she proceeded in applying them. This question mainly aims to promote reflection and the transfer of the training content to the school setting, in contrast to the standardized items we used for evaluation. In this study, we do not analyze the open-ended answers.

Each diary item is formulated as a state, meaning that the items ask for what the teacher did and thought on one particular day. When constructing the diary, we made sure to keep it at a suitable length for teachers to work on it regularly. We decided to make it no longer than one sheet of paper, which could be filled out in three to five minutes. For that reason, as well as for reasons of economy and evaluation, we chose the semi-standardized approach, with most of the items being standardized and just the one additional reflection item with an open answering format at the end. As a further source of motivation, teachers who completed at least 80 % of the diary entries were given the opportunity to get additional further education credit points, which German teachers needed to collect at the time the study took place.

Table 2 gives an overview of the standardized diary items with respect to the model content.

Table 2: Overview of the diary items

Number	Variable in the model	Item
1	Act systematically	Today, I proceeded systematically when assessing my pupils' learning behavior by considering which phase of the diagnostic process I am in.
2	Judgment formation	Today, I explicitly paid attention to specific judgment errors so that they do not bias my assessment.
3	Aim to foster (inverted)	Today, I assessed my students solely to grade them.
4	Aim to watch process	Today, to judge my pupils' learning behavior adequately, I compared their current learning behavior with their earlier learning behavior.
5	Make predictions	Today, I compared my prediction for one of my pupils with the learning behavior shown today in order to correct my impression if necessary.
6	Gather information	To find causes of learning difficulties for one of my pupils, I collected information from different sources today.
7	Methods	Today, in order to assess my pupils, I used methods apart from the usual examinations, such as observation sheets, pupils' self-assessments, or exchanges with colleagues.
8	Quality criteria	Today, I explicitly paid attention to the objectivity and reliability of my assessment methods.
9	Plan promotion	Today, I considered how to write a plan to promote one of my pupils' achievement.
10	Give feedback	Today, I gave feedback to a pupil or one of his parents on his learning behavior in a constructive way.
11	Teach SRL	In addition to normal class, I taught learning strategies today.

4. Results

4.1 Pretest-posttest comparison

There were no significant pretest differences in the dependent measures for the case scenario. Thus, the dependent measures were analyzed using a 2 (before/after the intervention) x 3 (training conditions) factorial ANOVA with time as a repeated measurement factor. Table 3 gives an overview of the results for the interaction *time x training condition* for the dependent measures in the case scenario. Means and standard deviations for the dependent measures are shown in Table 4.

Table 3: Overview of the results: ANOVA with repeated measures

Independent variable	Dependent variables	*df*	*F*	η^2	*f* [a]
Time x training condition	Scenario test: overall score	2/44	12.43***	.361	1.04
	Scenario test: preactional	2/44	5.48**	.199	1.06
	Scenario test: actional	2/44	6.37**	.224	.94
	Scenario test: postactional	2/44	2.77+	.112	.98

[a] $\sqrt{(QS_{time}/QS_{time\ x\ training})}$
+ $p < .10$; ** $p < .01$; *** $p < .001$.

Table 4: Means and standard deviations of the dependent measures on the pretest and posttest

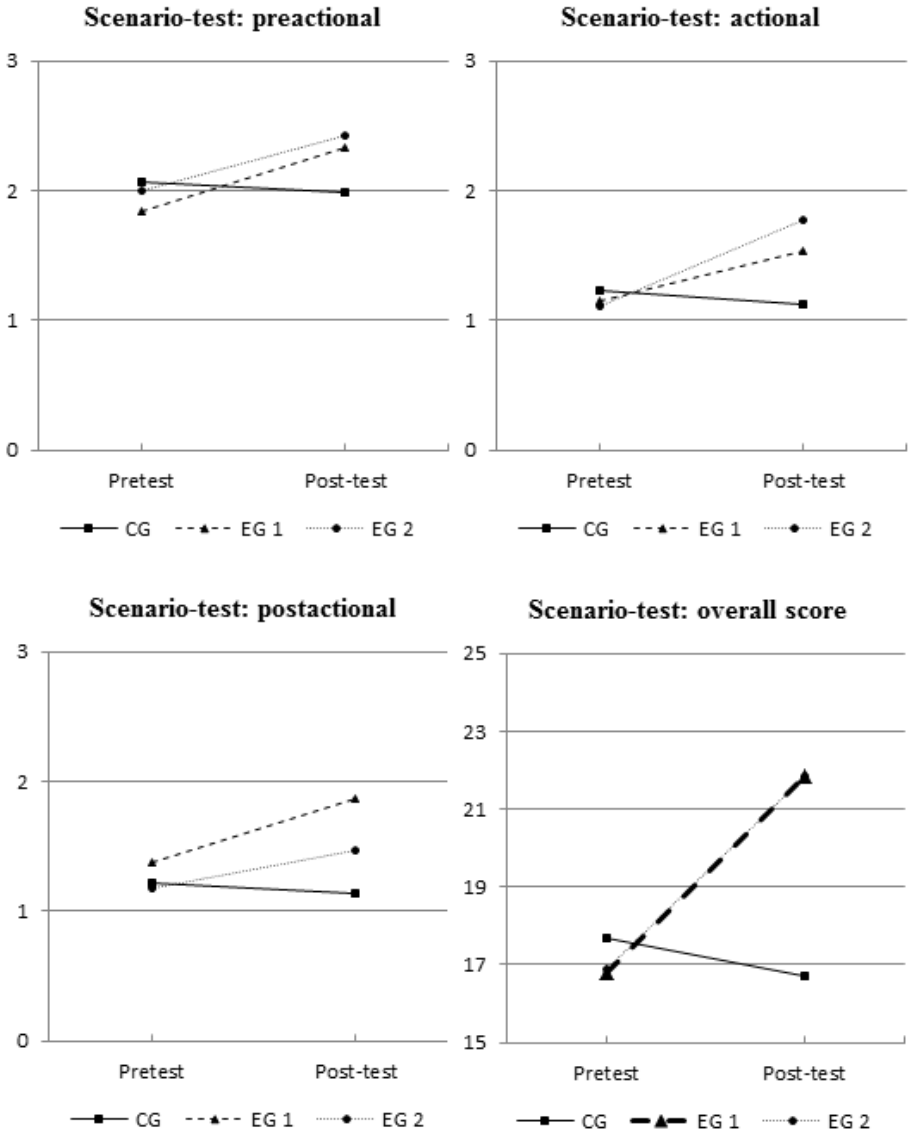
Group		EG 2 (training program & diary)		EG 1 (training program)		CG	
		M	*SD*	*M*	*SD*	*M*	*SD*
Scenario test: overall score	Pretest	16.87	5.40	16.80	5.07	17.71	3.29
	Posttest	21.87	4.02	21.87	5.18	16.71	4.14
Scenario test: preactional	Pretest	2.00	0.67	1.84	0.57	2.07	0.32
	Posttest	2.43	0.39	2.33	0.52	1.99	0.45
Scenario test: actional	Pretest	1.11	0.47	1.16	0.68	1.24	0.70
	Posttest	1.78	0.37	1.53	0.52	1.12	0.68
Scenario test: postactional	Pretest	1.18	0.58	1.38	0.84	1.22	0.58
	Posttest	1.47	0.63	1.87	0.71	1.14	0.55

Note. CG = control group.

As Table 3 shows, the interaction *time x training condition* is statistically significant for nearly every dependent variable with very high effect sizes. Only the postactional phase did not exhibit a statistically significant interaction, but there is a

tendency in the expected direction. Figure 2 illustrates these results for overall score in the scenario test, as well as for each of the three phases.

Figure 2: Results of the ANOVA with repeated measures



The figure shows that EG 1 and EG 2 improve by a comparable amount from pretest to posttest in their overall diagnostic competence values and in their scores for each phase, whereas the control group does not. It reveals that for all dependent variables, the statistically significant differences are due to differences between the experimental groups and the control group. Both experimental groups have a high increase in contrast to the control group, but the EG 1's increase does not differ from EG 2's increase.

Concerning the school subjects the participants teach, there is no difference in the pretest values of their overall diagnostic competence as well as in the improvement from pre- to posttest between math/natural sciences and language/social sciences teachers.

4.2 Process evaluation

In addition to the pretest-posttest comparison, the standardized diary items were used to perform time-series analyses. We computed trend analyses and interrupted time-series analyses based on the diary data from EG 2.

One hundred seventy-six of the 300 (59 % return rate) distributed diaries were included in the analyses. Trend analyses showed significant linear trends for most of the diary variables. Table 5 gives an overview of the linear trends on the item level for each phase.

Table 5: Linear trends of diary items

Phase	Item	*df*	*F*	*b*
Preactional	Aim to foster	1/15	7.22*	.04
	Aim to watch process	1/15	14.63**	.06
	Methods	1/15	0.24	.01
	Judgment formation	1/15	16.58**	.11
	Quality criteria	1/15	22.91***	.10
Actional	Make predictions	1/15	57.44***	.10
	Gather information	1/15	2.34	.04
	Act systematically	1/15	13.29**	.09
Postactional	Give feedback	1/15	0.37	.01
	Plan promotion	1/15	7.22*	.06
	Teach SRL	1/15	3.22	-.04

* $p < .05$; ** $p < .01$; *** $p < .001$.

Figure 3: Linear trends for the variables make predictions and plan promotion

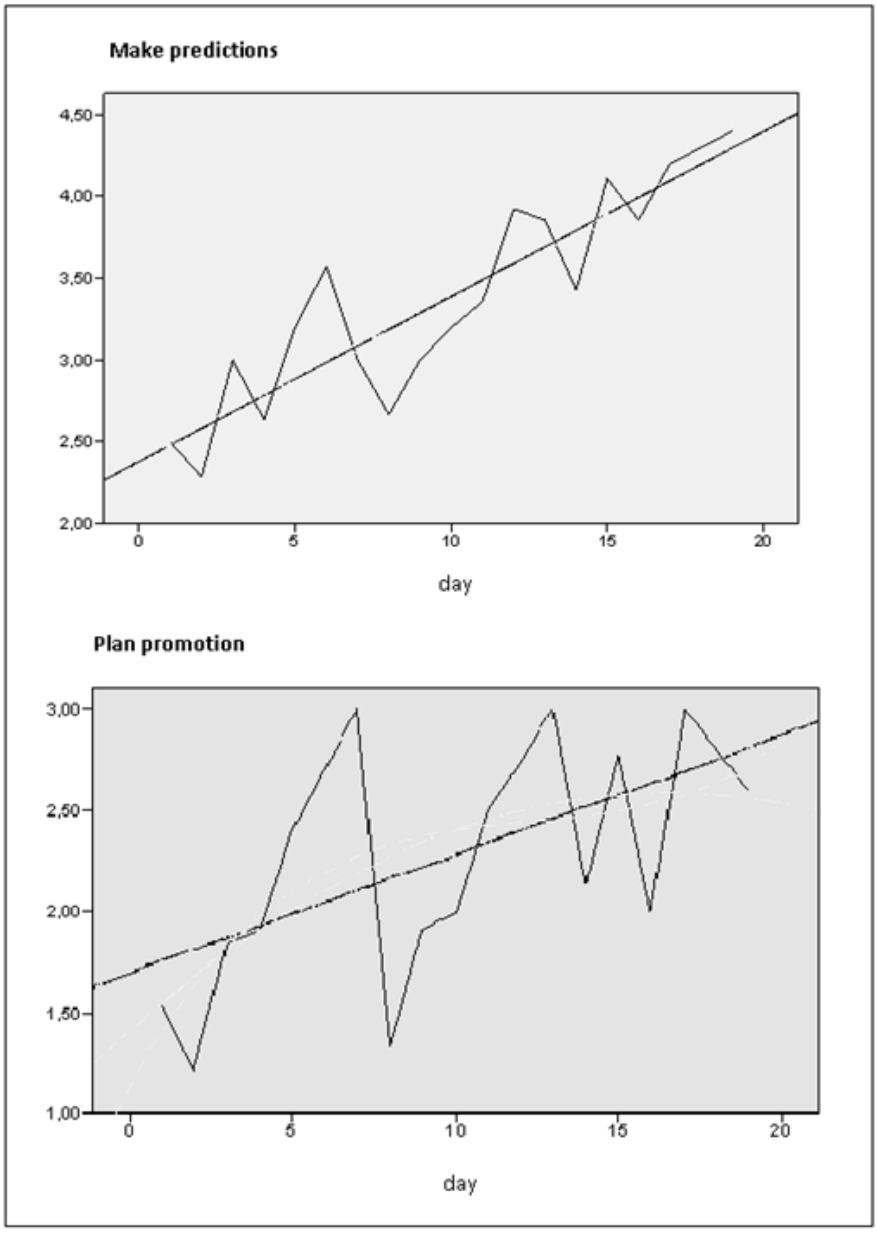


Figure 3 illustrates the statistically significant linear trends for the variables *make predictions* and *plan promotion*. The figure shows that in the course of the training program and the work on the diary, teachers made increasingly more predictions and put more and more effort into promoting students, a trend that continued one week after the last training session took place.

Interrupted time-series analysis is a well-known procedure for testing intervention effects (Campbell & Stanley, 1963). To conduct this method of analysis, the variables of interest need to be measured for a certain time before the intervention starts (baseline phase). After this phase, the intervention starts and the measurement points continue (intervention phase). By means of interrupted time-series analysis, the baseline and intervention phase can be tested for significant differences (Schmitz, Klug, & Schmidt, 2011). In this study, we implemented the baseline by letting the teachers work on the diary one week before the training program. With the help of interrupted time-series analyses, we analyzed the effect of each trained variable just after the training session in which that particular content was taught. Furthermore, we looked at the stability of the effects during the following week. The results show that there is an intervention effect in diary data that remains stable or increases for most of the preaction and action variables of diagnostic competence, but not for postaction variables. Table 6 gives an overview of the intervention effects of each trained variable and assigns the variables to the training session that dealt with them.

Table 6: Intervention effects of trained variables assigned to the corresponding training session

Phase	Item	Session	*t*	β
	Aim to foster	1	1.30	.31
	Aim to watch process	2	4.28**	.72
Preaction	Methods	2	1.49	.35
	Judgment formation	1	5.45***	.81
	Quality criteria	2	3.77**	.69
	Make predictions	2	5.57***	.81
Action	Gather information	2	1.34	.32
	Act systematically	1	3.54**	.66
	Give feedback	3	-0.79	-.19
Postaction	Plan promotion	3	0.91	.22
	Teach SRL	3	-0.47	-.12

** $p < .01$; *** $p < .001$.

Figure 4: Interrupted time-series analyses for the variables aim to watch process (Session 2) and judgment formation (Session 1)

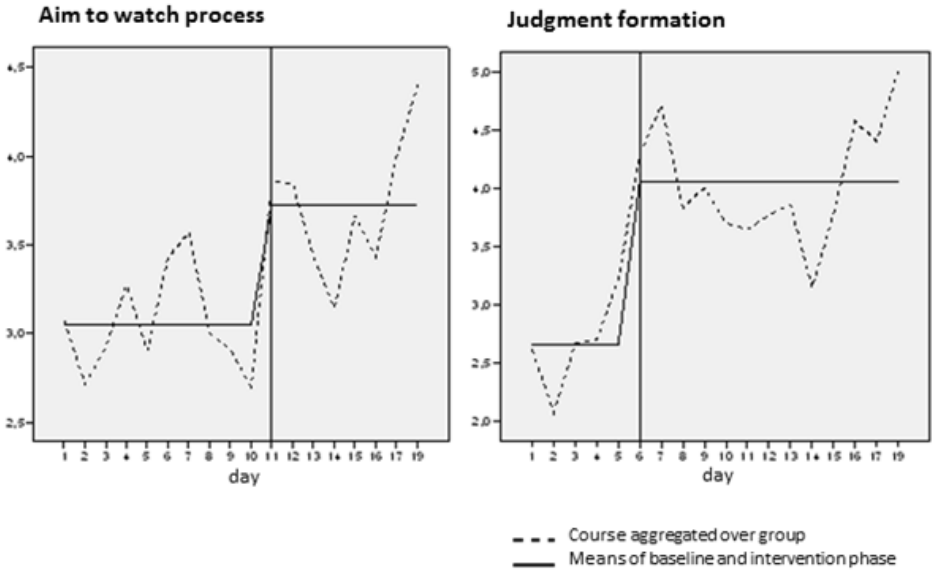


Figure 4 illustrates the intervention effects for the variables *aim to watch process*, which was covered in Session 2 (Day 11), and *judgment formation*, which was covered in Session 1 (Day 6). The figure illustrates that the mean for the baseline phase (until Day 11) is significantly lower than the mean after the variable has been trained. Looking at the dashed line, a further augmentation of the values can be recorded for some time after the intervention. The same pattern can be seen for the variable *judgment formation*, except that this variable was trained on Day 6.

5. Discussion

The pretest-posttest comparison showed that teachers benefit from the training program in comparison to a control group. Diagnostic competence increased in both training groups as reflected by the overall score and the scores for the preactional and actional phase. Effect sizes (see Table 3) were all very high, especially considering the short duration of the training program with only three 180-minute sessions.

For the postaction measures of diagnostic competence, the training program had no significant effect, but in the pretest-posttest comparison, there was at least a tendency in the expected direction. We suppose that this is due to the short duration of the training relative to the complexity of the content for the postaction phase. Planning the promotion of students, giving feedback to students and par-

ents, and teaching self-regulated learning is far too much content to be covered in one training session. Thus, we suggest that further studies broaden the postaction content (e.g., in the course of a modularization of the training program with one module for each phase consisting of several sessions). With more time and more practice with the postaction content, we expect significant effects to occur.

Concerning the diary, the return rate of 59 % can be considered a great success. Teachers accepted the short semi-standardized form of the diary and worked on it consistently over the four weeks. The standardization of most of the items and the short length of one page seem to be about right to motivate teachers to work on it. If the diary were longer and took more effort, the return rate would probably be lower. The use of incentives seems to be another important way of ensuring participants' commitment. Nevertheless, although the short, semi-standardized version of the diary might be good for measurement, analyses, and motivation, it is at the expense of teachers' reflection and the desired additional intervention effect. Reflection on experiences is seen as an important condition for the development of competences in expertise research (e.g., Strasser & Gruber, 2003). However, with just one open-ended question at the end of the diary, the expected additional intervention effect did not occur. Other than the short, semi-standardized form of the diary, a further possible explanation may be that a large amount of reflection was already integrated into and stimulated by the training program, so that teachers who did not work on the diary benefitted from self-monitoring as well. Maybe it was too much to expect that a short, semi-standardized version of a diary would offer an intervention effect above and beyond the training. Further studies should investigate whether a longer diary with more open-ended reflection questions would lead to an intervention effect. Additionally, open-ended questions could provide ideas for further research. However, the semi-standardized diary proved to be a helpful instrument for measuring the transfer of the trained variables to teachers' everyday work at school. With the diary, teachers reflected upon which training content they actually implemented in their classes every day. Furthermore, the process data collected by the diaries gave us much more insight into which diagnostic actions teachers applied over time, and how this augmentation of competence functioned. We expected positive linear trends over the four weeks. We found these trends for seven of the eleven measured variables. There was even a significant linear trend for the postactional variable *plan promotion*, which we did not find in the pretest-posttest comparison, even when we analysed it on the item level. Thus, the diary data provided us with additional information. We also obtained additional information by looking at the interrupted time-series analyses. For many of the trained variables, we were able to illustrate and support an augmentation of the scores from the baseline level just after the session in which the specific variable was trained. Furthermore, we were able to see whether the augmentation persisted, declined, or increased even further by looking at the progression of the variable over time. For the five significant variables, scores rose further, indicating a long-term and even delayed effect of the training program and the work on the diary. However, in further studies, implementing a follow-up test to complement

the pretest-posttest comparison would be great as an additional verification of the lasting effects measured by the other instruments. Postactional variables exhibited no intervention effect in the process data. In addition to the necessity of extending the training program, especially when it comes to postactional content, we assume that the operationalization of the postactional variables in the diary needs to be improved. Teachers do not and cannot implement these variables (plan promotion, give feedback, teach SRL) every day at school. Thus, the formulation of the items which, for example, ask whether the teacher taught SRL that day, cannot be answered positively every day, even if the intervention is effective. The postactional diary items need to be reformulated, perhaps asking whether the teacher gave thought to those variables on that particular day. Further studies should address whether interrupted time-series analyses will show effects for postactional variables if they are reformulated in this way.

In contrast to the scenario-test, the diary is a self-report measure which is prone to the usual limitations and biases that can occur in self-reports – e.g., socially desired answers. Adding, for example, the students' perspective on what the teacher does in the classroom and whether change was noticed could help here. In summary, the multi-method approach we chose in this study is an advantage and offers great potential. For further studies, additional measures with high ecological validity such as real classroom observations or the analyses of video vignettes should be implemented, and convergent validity among the measures should be tested.

In this study, we chose secondary school teachers from a German *Gymnasium* (grammar school) and comprehensive school as the sample. Future research should aim to optimize and evaluate the training program and the diary for other school forms and test whether they are generalizable.

The study design was quasi-experimental, with the advantages of being longitudinal, combining pretest and posttest measures with process data, and having a control group. However, the main disadvantage is that in the field, no randomized assignment of the teachers to the conditions was possible because the teachers were trained in their schools and had limited time for the appointments. But we did randomly assign which group of participants would be in the experimental and which in the control conditions. Furthermore, in each of the three training groups, half of the participants worked on the diary. In sum, the quasi-experimental design suited the opportunities in the field.

6. Conclusion

The current study aimed to contribute to promoting teachers' diagnostic competence when it comes to diagnosing students' learning behavior. Both the procedure and the findings of this study provide an opportunity to draw several implications for European teacher education. As diagnosing is an everyday task at school,

and with pressures to provide individualized support to students increasing, such a training concept could be helpful both in initial teacher education and in further education.

Teacher educators and those involved in the professional development of teachers can make use of the model's theoretical framework, focusing their training courses on the model components and process character. Teaching programs can be enriched by discussing diagnosing on the basis of the model. Additionally, various tasks used in the training that proved to be helpful can be used in teacher education to promote teachers' diagnostic competence. Case scenarios such as those used in the training program can be used to practice diagnosing, following the model in a safe environment, and in discussion with fellow students, who can benefit from each other, and the feedback given by teacher educators. In addition to the training content, the diary can also be implemented in teacher education to help teachers reflect upon their diagnostic action. It can also be used to make small improvements in diagnosing visible, which can in turn be a source of motivation for teachers. However, to use it genuinely for education and further education, the diary should be adapted or further developed in terms of including more open-ended questions to strengthen the reflective element. Teachers may learn from reflecting upon their diagnostic action and may be able to adapt their diagnosing and teaching in class. Thus, both teachers and students can benefit.

In conclusion, the aim should be to increase knowledge and competence in this area by continuing to train teachers and, most importantly, by implementing this concept or aspects of it in teacher education.

Acknowledgements

The preparation of this paper was supported by the grants SEHM 1538/5-2 from the German Research Foundation (DFG) in the priority program *Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes* (SPP 1293).

References

Abs, H. J. (2007). Überlegungen zur Modellierung diagnostischer Kompetenz bei Lehrerinnen und Lehrern [Considerations on modelling teachers' diagnostic competence]. In M. Lüders & J. Wissinger (Eds.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* [Research of teacher education. Competence development and programme evaluation] (pp. 63–84). Münster, Germany: Waxmann.

Bakkenes, I., Vermunt, J. D., & Wubbels, T. (2010). Teacher learning in the context of educational innovation: Learning activities and learning outcomes of experienced teachers. *Learning and Instruction, 20*(6), 533–548. doi: 10.1016/j.learninstruc.2009. 09.001

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*(5), 463–482. doi: 10.1016/S0883-0355(02)00004-6

Brante, G. (2009). Multitasking and synchronous work: Complexities in teacher work. *Teaching and Teacher Education, 25*(3), 430–436. doi: 10.1016/j.tate.2008.09.015

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. doi: 10.3102/00346543065003245

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*(8), 947–967. doi: 10.1016/S0742-051X(02)00053-7

Coladarci, T. (1986). Accuracy of teacher Judgments of student responses to standardized test items. *Journal of Educational Psychology, 78*(2), 141–146.

Dignath, C., Büttner, G., & Langfeldt, H. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Psychology Review, 3*(2), 101–129.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *American Medical Association, 287*(2), 226–237. doi:10.1001/jama.287.2.226

Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly, 18*(1), 52–65. doi: 0.1521/scpq.18.1.52.20876

Fiedler, K., Walther, E., Freytag, P., & Plessner, H. (2002). Judgment biases in a simulated classroom – A cognitive-environmental approach. *Organizational Behavior and Human Decision Processes, 88*(1), 527–561.

Hattie, J., Biggs, H., & Purdie, N. (1996) Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*(2), 99–136.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. doi: 10.3102/003465430298487

Helmke, A., Hosenfeld, I., & Schrader, F. W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften [Comparative tests as an instrument for the improvement of teachers' diagnosic competence]. In R. Arnold & C. Griese (Eds.), *Schulmanagement und Schulentwicklung* [School management and development] (pp. 119–144). Hohengehren, Germany: Schneider.

Helmke, A., & Schrader, F. W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teaching & Teacher Education, 3*(2), 91–98. doi: 10.1016/0742-051X(87)90010-2

Horstkemper, M. (2004). Diagnosekompetenz als Teil pädagogischer Professionalität [Diagnosis competence as a part of educational professionalism]. *Neue Sammlung, 44*(2), 201–214.

Jäger, R. S. (2007). *Beobachten, bewerten, fördern. Lehrbuch für die Aus-, Fort- und Weiterbildung* [Observe, value, foster. Textbook for education and further education]. Landau, Germany: Empirische Pädagogik.

Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E., & Vollmer, H. J. (2003). *Expertise zur Entwicklung nationaler Bildungsstandards* [Expertise for the development of national educational standards]. Berlin, Germany: Bundesministerium für Bildung und Forschung.

Klug, J., Bruder, S., & Schmitz, B. (2016). Which variables predict teachers diagnostic competence when diagnosing students' learning behavior at different stages of a teachers' career? *Teachers and Teaching: Theory and Practice, 22*(4), 461–484.

Klug, J., Bruder, S., Kelava, A., Spiel, C., & Schmitz, B. (2013). Diagnostic competence of teachers: A process model that accounts for diagnosing learning behavior tested by means of a case scenario. *Teaching and Teacher Education, 30*, 38–46. doi: 10.1016/j.tate.2012.10.004

Klug, J., Bruder, S., Keller, S., & Schmitz, B. (2012). Hängen diagnostische Kompetenz und Beratungskompetenz von Lehrkräften zusammen? – Eine korrelative Untersuchung [Are diagnostic and consulting competencies of teachers related? A correlational study]. *Psychologische Rundschau, 63*(1), 3–10.

Klug, J., Krause, N., Schober, B., Finsterwald, M., & Spiel, C. (2014). How do teachers promote their students' lifelong learning in class? Development and first application of the LLL-Interview. *Teaching and Teacher Education, 37*, 119–129. doi: 10.1016/j.tate.2013.09.004

Kukla-Acevedo, S. (2009). Do teacher characteristics matter? New results on the effects of teacher preparation on student achievement. *Economics of Education Review, 28*(1), 49–57. doi: 10.1016/j.econedurev.2007.10.007

Lan, W. Y. (1996). The effects of self-monitoring on students' course performance, use of learning strategies, attitude, self-judgement ability, and knowledge representation. *The Journal of Experimental Education, 64*(2), 101–115.

Lee, K., Chiu, S. N., Hasselt, C. A. van, & Tong, M. (2009). The accuracy of parent and teacher reports in assessing the vocabulary knowledge of Chinese children with hearing impairment. *Language, Speech, and Hearing Services in Schools, 40*(1), 31–45.

Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish-little-pond effect. *Contemporary Educational Psychology, 30*(3), 263–285. doi: 10.1016/j.cedpsych.2004.10.002

Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualisations. *Teaching and Teacher Education, 20*(5), 523–535. doi: 10.1016/j.tate.2004.04.008

Marcos, J. J. M., Miguel, E. S., & Tillema, H. (2009). Teacher reflection on action: What is said (in research) and what is done (in teaching). *Reflective Practice, 10*(2), 191–204. doi: 10.1080/14623940902786206

McGinn, T., Jervis, R., Wisnivesky, J., Keitz, S., & Wyer, P. C. (2008). Tips for teachers of evidence-based medicine: Clinical Prediction Rules (CPRs) and estimating pretest probability. *Journal of General Internal Medicine, 23*(8), 1261–1268.

Opdenakker, M. C., & Van Damme, J. (2006). Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. *Teaching and Teacher Education, 22*(1), 1–21. doi: 10.1016/j.tate.2005.07.008

Ophuysen, S. van (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnempfehlung [Comparison of diagnostic decisions of novices and experts using the example of school career recommendations]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 38*(4), 154–161. doi: 10.1026/0049-8637.38.4.154

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: When measuring and intervening come hand in hand. *Scandinavian Journal of Educational Research, 60*(6), 723–735.

Perels, F., Dignath, C., & Schmitz, B. (2009). Is it possible to improve mathematical achievement by means of self-regulation strategies? Evaluation of an intervention in regular math classes. *European Journal of Psychology of Education, 24*(1), 17–32.

Perry, N. E., Hutchinson, L., & Thauberger, C. (2008). Talking about teaching self-regulated learning: Scaffolding student teachers' development and use of practices that promote self-regulated learning. *International Journal of Educational Research, 47*(2), 97–108.

Pintrich, P. R., & de Groot, E. van (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–50. doi: 10.1037/0022-0663.82.1.33

Schmitz, B. (2006). Advantages of studying processes in educational research. *Learning and Instruction, 16*(5), 433–449.

Schmitz, B., Klug, J., & Schmidt, M. (2011). Assessing self-regulated learning using diary measures with university students. In B. Zimmerman & D. Schunck (Eds.), *Handbook of self-regulation of learning and performance* (pp. 251–266). New York, NY: Routledge.

Schmitz, P., & Perels, F. (2011). Self-monitoring of self-regulation during math homework behaviour using standardized diaries. *Metacognition and Learning, 6*(3), 255–273. doi: 10.1007/s11409-011-9076-6

Schmitz, B., & Wiese, B. (2006). New perspectives for the evaluation of training sessions in self-regulated learning: Time-Series analyses of diary data. *Contemporary Educational Psychology, 31*(1), 64–96.

Sowa, P. A. (2009). Understanding our learners and developing reflective practice: Conducting action research with English language learners. *Teaching and Teacher Education, 25*(8), 1026–1032. doi: 10.1016/j.tate.2009.04.008

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz [Accuracy of teachers' judgment of pupils' characteristics and the construct of diagnostic competence]. *Zeitschrift für Pädagogische Psychologie, 19*(1/2), 85–95. doi: 10.1024/1010-0652.19.1.85

Strasser, J., & Gruber, H. (2003). Kompetenzerwerb in der Beratung: Eine kritische Analyse des Forschungsstands [Competence acquisition in consultation: A critical analysis of the state of research]. *Psychologie in Erziehung und Unterricht, 50*(4), 381–399.

Vogt, F., & Rogalla, M. (2009). Developing adaptive teaching competency through coaching. *Teaching and Teacher Education, 25*(8), 1051–1060. doi: 10.1016/j.tate.2009.04.002

Wang, M. C. (1973). The accuracy of teacher's predictions on children's learning performance. *The Journal of Educational Research, 66*(10), 462–465.

Webber, J., Scheuermann, B., McCall, C., & Coleman, M. (1993). Research on self-monitoring as behavior management technique in special education classrooms: A descriptive review. *Remedial and Social Education, 14*(2), 38–56. doi: 10.1177/074193259301400206

Wilson, E. B. (1952). *An introduction to scientific research*. New York, NY: McGraw-Hill.

Wirth, J., & Leutner, D. (2008). Self-regulated learning as a competence: Implications of theoretical models for assessment methods. *Zeitschrift für Psychologie, 216*(2), 102–110. doi: 10.1027/0044-3409.216.2.102

Yen, C., Konold, T. R., & McDermott, P. A. (2004) Does learning behavior augment cognitive ability as an indicator of academic achievement? *Journal of School Psychology, 42*(2), 157–169.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–35), San Diego, CA: Academic Press.